ELEC 476/576: Introduction to Machine Learning

Instructor: Reinhard Heckel rh43@rice.edu, office number: Duncan 2048, office hours: Wednesday 1pm-2:30pm

Teaching assistants: Daniel LeJeune dlejeune@rice.edu, Lorenzo Luzi ll65@rice.edu, office hours every Thursday, 4:30pm-5:45pm in Herzstein 210

Time and Location: Tuesday and Wednesday 4:30pm-5:45pm in Herzstein 210

Catalogue description: The course provides an introduction to concepts, methods, best practices, and theoretical foundations of machine learning. Topics covered include regression, classification, model selection and validation, kernels, nearest neighbor algorithms, clustering, decision trees, ensemble learning, empirical risk minimization and regularization, learning theory, unsupervised learning, and neural networks.

Learning outcomes: Students will learn the theoretical foundations of machine learning algorithms, how and when to apply those algorithms, their comparative strengths and weaknesses, and how to critically evaluate their performance. Students completing this course should be able to (i) apply basic machine learning methods to build predictive models or perform exploratory analysis, (ii) properly tune, select and validate machine learning models, (iii) interpret their results, and (iv) understand their limits.

List of required texts: We do not follows a textbook, lecture notes will be distributed. Helpful references include:

- "Elements of Statistical Learning" by Hastie, Tibshirani & Friedman
- "Machine Learning" by Tom Mitchell
- "Foundation of Machine Learning", by Mohri, Rostamizadeh, and Talwalkar,
- "Understanding Machine Learning: From Theory to Algorithms" by Shalev-Shwartz and Ben-David.

Grade policies: Homeworks 40%, First exam on October 18, counts 25%, second exam on November 30, counts 35%. Both exams will be take-home exams distributed electronically on the specified day and to be completed during a 24 hour period for 3-4 hours.

Prerequisites:

- Basic computer science skills, at a level to perform coding assignments in a language of choice (Python or R are recommended, but Matlab and other languages are equally fine) for example Stat 405, Caam 210, Comp 140, or Comp 130.
- Basic linear algebra, for example Caam 335 or Math 355.
- Basic probability theory and statistics:
- Helpful but not required are introductory courses to convex optimization, signal processing, statistics (e.g., Stat 411), and learning (Elec 301 and Elec 475/575).

Course overview: Here is a tentative syllabus, that is subject to change:

- Lecture 1: Overview of the course
- Lecture 2: Linear regression
- Lecture 3: Ridge regression
- Lecture 4: Model selection an validation
- Lecture 5-6: Convex optimization
- Lecture 7-8: Logistic regression, generative models and naive Bayes
- Lecture 9-10: Support vector machines and kernels
- Lecture 11: The stochastic gradient method
- Lecture 12: Nearest neighbor algorithms
- Lecture 13-14: Learning theory (bias/variance tradeoff; overfitting; generalization error)
- Lecture 15: Decision trees, bagging, random forests
- Lecture 16: Strong and weak learning, boosting
- Lecture 17: k-means- and spectral clustering
- Lecture 18: Mixture of Gaussians and the EM algorithm
- Lecture 19: Principal component analysis
- Lecture 20: Factor analysis (e.g., matrix factorization)
- Lecture 21-24: Neural networks, deep learning, convolutional neural networks
- Lecture 25-26: Active learning

Homework Policy: There are seven bi-weekly homeworks. Homeworks are submitted online via www.gradescope.com, access code M6ZKDE. Each homework has core questions worth 20 points, and a 578-question worth 5 points. For students taking ELEC 478, only core questions are graded. For students taking ELEC 578, core questions and the 578-question are graded. Your weakest homework is dropped.

Homeworks must be submitted by the posted due date. There is no credit for late works, and there are no exceptions (unless based on university policies such as medical reasons). If you are not able to comply with the late homework policy due to travel plans or any other reason please do not enroll in the course. However, recall that your weakest homework is dropped which gives you some flexibility.

All grading related requests must be submitted to the TAs via email only—office hours and discussion sessions are for discussions and questions related to the course material only.

Discussions among students on the homeworks are allowed and encouraged, but you must hand in your own solutions and write and run your own code. In addition, you must indicate who you collaborated with.

We sometimes pose questions that can be answered by looking up solutions in books or webpages. We expect students not to copy solutions if they exist from any source. The homework is to help you to learn the material, and we expect you to make a serious effort in solving the problems. If you use other material, it must be acknowledged with citation.

Attendance Policy: Student are encouraged, but not required to attend lectures.

Rice Honor Code: In this course, all students will be held to the standards of the Rice Honor Code, a code that you pledged to honor when you matriculated at this institution. If you are unfamiliar with the details of this code and how it is administered, you should consult the Honor

System Handbook at http://honor.rice.edu/honor-system-handbook/. This handbook outlines the University's expectations for the integrity of your academic work, the procedures for resolving alleged violations of those expectations, and the rights and responsibilities of students and faculty members throughout the process.

Disability Support Services: If you have a documented disability or other condition that may affect academic performance you should: 1) make sure this documentation is on file with Disability Support Services (Allen Center, Room 111 / adarice@rice.edu / x5841) to determine the accommodations you need; and 2) talk with me to discuss your accommodation needs.

Syllabus Change Policy: This syllabus is only a guide for the course and is subject to change with advanced notice.